AI FRAMEWORK FOR CRIME ANALYSIS

Dr.S.N.Zaware[1], , Monica Gupta , Preeti Ankam,Aakanksha Dumane, Adhiraj Budukh
[1]HOD of Computer Dept., Department of Computer Engineering
AISSMS Institute Of Information Technology ,Kennedy rd., Pune,411001, India
Department of Computer Engineering AISSMS Institute Of Information Technology ,
Kennedy rd., Pune,411001, India

Received: February 11, 2019 Accepted: March 16, 2019

ABSTRACT: *Crime analysis is a function of identifying the patterns and analyzing the rapidly increasing crime trends. It helps in highlighting the events and incidents that may need further investigation. This information helps police departments and other crime departments to solve the crimes faster and effectively. The challenge faced by police departments is difficulty in analyzing the large volume of crime data and criminal behaviour. In order to identify the patterns in crime Data mining plays an important role. This paper focuses on fetching real time data from RSS Feeds. This data will be further divided into crime and non-crime news. The crime news will be categorized into various types like law and order, financial and non financial crimes etc. Finally a graphical representation of the most emerging crimes will be shown.*

Key Words: *Rss feeds, crime analysis, classification, clustering*

I. INTRODUCTION

Now-a-days, there has been a fringe rise in crimes in India. Fierce crimes, Property crimes, organized crime, the illegal drug trade, arms trafficking, degeneracy of lifestyle are the most common types of crimes occurring today. As the occurrence of crimes is unpredictable, it is hard to gain control over them for local police officers. It is difficult to analyze the victims of crime but the place where crime has been occurred or happened can beanalyzed. So, there is a need of an effective analyzing tool which can analyze crime data efficiently and quickly to give some useful crime patterns or predict future crimes. This can prove helpful to the intelligence agencies or local law enforcement agencies.

In this paper, Data from different RSS Feeds is obtained and applied various data mining techniques on it to gain interesting patterns on crime rates. Classification is asupervised class prediction technique. Data in the data warehouse is filtered and using classification algorithms , the crime data is selected to actupon. The Naïve Bayes classification algorithm is applied on real time data. This classifiesthe data into various categories of crimes.

A number of clustering techniques are used in data mining. Using the suitable clustering technique, the classified data is clustered and a final output in the form of graph is displayed.

II. LITERATURE REVIEW

According to our survey, crime analysis was performed on past records.Criminal investigation analysis (CIA) tool is also used to solve violent crimes and the accuracy was limited. The forensic tool kit generates the file and analyzes the data and also used to analyze the victim. Data mining techniques like K-Means, Influenced Association Classifier and J48 Prediction tree were also used for investigating the cyber crime data and sorts out the problem.

Supervised, semi-supervised and unsupervised learning techniques are carried out on the crime records for knowledge discovery and to help in increasing the predictive accuracy of the crime. It is also concluded that classification techniques gave more than 90% accuracy using Bayes theorem.

K.ChitraLekha and Dr. S. Prakasam implemented data mining techniques like K-Means, Influenced Association Classifier and J48 Prediction tree for investigating the cyber crime data and sorted out the problem.

Chhaya Chauhan and Smriti Sehgal carried out a review on various classification techniques in data mining. They concluded that classification techniques gave more than 90% accuracy using Bayes theorem.

III. ARCHITECTURE DIAGRAM

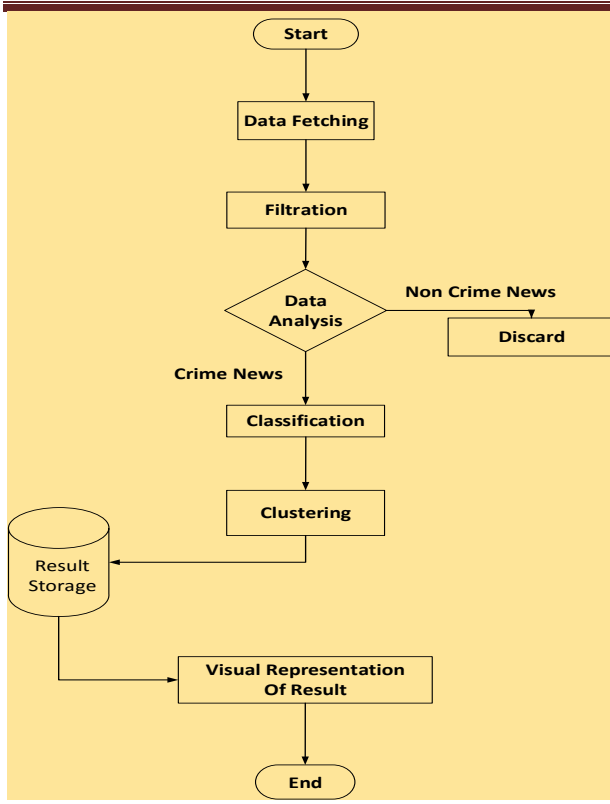

Fig.1

IV. PROCESSING STEPS

I. DATA ACQUISITION AND FILTERING

This step performs approximation algorithm. This is done by obtaining live RSS Feed Data and then filtering and dividing them into segments.

The input is Live Data Feed which is the processed data set and the output is filtered data in key value pair . This output is sent to processing Mechanism. Here in step 1, related details are filtered out.

In step 2, filtered data are the association of different key value pairs and each pair is different numbers of sample, which results in forming a data block.

1. Retrieve the data from RSS Feeds and filter the related data i.e. processed data. All other unnecessary data will be discarded.
2. The processed data is divided into suitable or proper key-value pair.
3. The unprocessed data is directly transmitted to aggregation step without processing.
4. Each distinct data block of processed data is allocated and transmitted to various processing steps in Data Processing Unit.

RSS FEED: It contains frequently updated news headlines in xml format. This data is available to users who want to receive timely updates from website or to aggregate data from many different sites.

MEANING CLOUD: It is a software as a service product used for text analysis and semantic processing. The different API's available are:-topic extraction, text classification, sentiment analysis, etc.

II. PROCESSING AND CALCULATION

In this step , the data blocks are forwarded to beprocessed by Data Processing Unit.

The input to this step is Filtered Data and the output of this step will be Normalized News data into Numerical comparable form Along with Historical Values.

1. For each event data, relevant Historical Data is extracted.
2. Normalize this for all the live feed.
3. Persist the data into data store and forward it.

III.MATHEMATICAL MODEL

1) We calculate the probability of news in each news source

$$\pi_j = \frac{class_j}{\sum_{n=1}^{20} class_n}$$

2) Probability of each word per class:-
Here we will find average of each word for a given class

For class j and word i ,

$$P(i|j) = \frac{word_{ij}}{word_j}$$

3) Since some words will have 0 count , we will perform Laplace smoothing

$$P(i|j) = \frac{word_{ij} + \alpha}{word_j + |V| + 1},\ \alpha = 0.001$$

V is the array of all words in vocabulary.

4) Also, in order to take stop words into account, we will add a Inverse Document Frequency (IDF)weight on each word:

$$t_i = \log\left(\frac{\sum_{n=1}^{N} doc_n}{doc_i}\right)$$

IDF is intended to reflect how important a word is to a document in a collection.

IV. APPLICATIONS

Cultivating people about information of crime. It is important for public concerns and in public action towards crime.

Public cognizant is also important in reducing the crimes in local areas.

Supporting trend and pattern analysis. Integrating all i.e. traditional and nontraditional data for analysis improvement

Prediction of occurrence of future crime. Possible to take Strategic decisions in rapidly changing environment.

Understand events and dynamics of crime hazards, persons, and events.

V. RESULTS

The result is displayed in the graphical form as shown below:

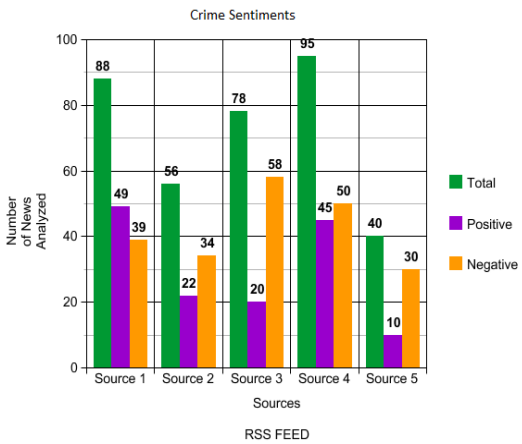

Fig. 2

VI. CONCLUSION

In this paper, we have applied classification and clustering algorithms to get the better results that can help police department, law enforcement officers to improve their work. It will increase the efficiency in solving the crimes faster.

VII. REFERENCES

1. ShijuSathyadevan, Devan M.S and SuryaGangadharanS,"Crime Analysis and Prediction Using Data Mining",2014
2. AbbaBabakura , Md Nasir Sulaiman and Mahmud A. Yusuf "Improved Methods of Classification algorithm for Crime Prediction",2014
3. Nisha and Puneet Jai Kaur, "A survey Of Clustering Techniques and Algorithms", 2015
4. UbonThongsatapornwatana ,"A Survey of Data Mining Techniques for Analyzing Crime Patterns",2016
5. NidhiTomar and Amit Kumar Manjhvar, "An Improved Optimized Clustering Technique For Crime Detection", 2016
6. Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav ,"Crime Pattern Detection, Analysis and Prediction ",2017
7. Adel Ali Alkhaibari, Ping-Tsai Chung, "Cluster Analysis for Reducing City Crime Rates ", 2017
8. Chhaya Chauhan and Smriti Sehgal , "A Review: Crime Analysis Using Data Mining Techniques And Algorithms.",2017
9. Priyanka Das and Asit Kumar das, "Crime Analysis against Women from Online Newspaper Reports and an Approach to apply it in Dynamic Environment.", 2017
10. K. Chitra Lekha and Dr. S. Prakasam, " Data Mining Techniques in Detecting andPredicting Cyber Crimes in Banking Sector ",2017